

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:

- (A) NAME: Kossmann, Jens
Buttcher, Volker
Welsh, Thomas

(ii) TITLE OF INVENTION: DNA SEQUENCES ENCODING ENZYMES CAPABLE OF FACILITATING THE SYNTHESIS OF LINEAR ALPHA-1,4 GLUCANS IN PLANTS, FUNGI AND MICROORGANISMS

(iii) NUMBER OF SEQUENCES: 4

(iv) CORRESPONDENCE ADDRESS:

- (A) ADDRESSEE: James F. Haley, Jr., c/o Fish & Neave
- (B) STREET: 1251 Avenue of the Americas
- (C) CITY: New York
- (D) STATE: New York
- (E) COUNTRY: United States of America
- (F) ZIP: 10020

(v) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Floppy disk
- (B) COMPUTER: IBM PC compatible
- (C) OPERATING SYSTEM: PC-DOS/MS-DOS
- (D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(vi) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER: DE P 44 17 879.4
- (B) FILING DATE: 18-MAY-1994

(vi) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER: DE P 44 47 388.5
- (B) FILING DATE: 22-DEC-1994

(vii) ATTORNEY/AGENT INFORMATION:

- (A) NAME: James F. Haley, Jr., Esq.
- (B) REGISTRATION NUMBER: 27,794
- (C) DOCKET NUMBER: GFB-1

(viii) TELECOMMUNICATION INFORMATION:

- (A) TELEPHONE: (212) 596-9000
- (B) TELEFAX: (212) 596-9090

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2914 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Neisseria polysaccharea

(vii) IMMEDIATE SOURCE:

(A) LIBRARY: genomic library in pBluescriptII SK
(B) CLONE: pNB2

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION: 957..2867

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

GAGTTTGCG	TTCCCGAAC	GAACGTGATG	CTTGAGCCGA	ACACCTGTCC	GGCAAGGC	60
CTGACCGCCC	CCTTTGCC	CATCGACATC	GTAACAATCG	GTTTGGTGGC	AAGCTCTTC	120
GCTTGAGCG	TGGCAGAAAG	CAAAGTCAGC	ACGTCTCCG	CGCTTGC	CATCACCGCA	180
ATTTGCAGA	TGTCCGCC	GCAGTCCTCC	ATCTGTTCA	GACGGCATA	C GATTCTTCT	240
TGCAGCGCG	TGCGGTGAAA	CTCATGATTG	CAGAGCAGGG	CGGCGATGCC	GTTTTTTGA	300
GCATGCGCCA	CGGCGCGCCG	GACGGCGGTT	TCGCCGGAAA	AAAGCTCGAT	ATCGATAATG	360
TCGGGCAGGC	GGCTTCAAT	CAGCGAGTCG	AGCAGTTCAA	AATAATAATC	GTCCGAACAC	420
GGGAACGAGC	CGCCTCGCC	ATGCCGTCTG	AACGTAAACA	GCAGCGGCTT	GTCGGCAGC	480
GGCGTCGCGA	CGGTCTCGT	GTGGCGCAAT	ACTTCGCCGA	TGCTGCCCGC	GCATTCCAAA	540
AAATCGGCGC	GGAACTCGAC	GATATCGAAG	GGCAGGTTTT	TGATTGGTC	AAGTACGGCG	600
GAAAGTACGG	CGGCATCGCG	GGCGACAAGC	GGCACGGCGA	TTTTGGTGC	TCCGCTTCCG	660
ATAACGGTGT	TTTGACGGT	CAGGCTGGTG	TGCATGGCG	TTGTTGCGGC	TGAAAGGAAC	720
GGTAAAGACG	CAATTATAGC	AAAGGCACAG	GCAATGTTTC	AGACGGCATT	TCTGTGCGGC	780
CGGCTTGATA	TGAATCAAGC	AGCATCCGCA	TATCGGAATG	CAGACTTGGC	ACAAGCCCTG	840
TCTTTCTAG	TCAGTCCGCA	GTTCTTGCAG	TATGATTGCA	CGACACGCC	TACACGGCAT	900
TTGCAGGATA	CGGCGGCAGA	CCGCCGGTCG	GAAACTTCAG	AATCGGAGCA	GGCATC	956
ATG TTG ACC CCC ACG CAG CAA GTC GGT TTG ATT TTA CAG TAC CTC AAA						1004
Met Leu Thr Pro Thr Gln Gln Val Gly Leu Ile Leu Gln Tyr Leu Lys						
1 5 10 15						
ACA CGC ATC TTG GAC ATC TAC ACG CCC GAA CAG CGC GCC GGC ATC GAA						1052
Thr Arg Ile Leu Asp Ile Tyr Thr Pro Glu Gln Arg Ala Gly Ile Glu						
20 25 30						
AAA TCC GAA GAC TGG CGG CAG TTT TCG CGC CGC ATG GAT ACG CAT TTC						1100

Lys Ser Glu Asp Trp Arg Gln Phe Ser Arg Arg Met Asp Thr His Phe			
35	40	45	
CCC AAA CTG ATG AAC GAA CTC GAC AGC GTG TAC GGC AAC AAC GAA GCC			1148
Pro Lys Leu Met Asn Glu Leu Asp Ser Val Tyr Gly Asn Asn Glu Ala			
50	55	60	
CTG CTG CCT ATG CTG GAA ATG CTG CTG GCG CAG GCA TGG CAA AGC TAT			1196
Leu Leu Pro Met Leu Glu Met Leu Leu Ala Gln Ala Trp Gln Ser Tyr			
65	70	75	80
TCC CAA CGC AAC TCA TCC TTA AAA GAT ATC GAT ATC GCG CGC GAA AAC			1244
Ser Gln Arg Asn Ser Ser Leu Lys Asp Ile Asp Ile Ala Arg Glu Asn			
85	90	95	
AAC CCC GAT TGG ATT TTG TCC AAC AAA CAA GTC GGC GGC GTG TGC TAC			1292
Asn Pro Asp Trp Ile Leu Ser Asn Lys Gln Val Gly Val Cys Tyr			
100	105	110	
GTT GAT TTG TTT GCC GGC GAT TTG AAG GGC TTG AAA GAT AAA ATT CCT			1340
Val Asp Leu Phe Ala Gly Asp Leu Lys Gly Leu Lys Asp Lys Ile Pro			
115	120	125	
TAT TTT CAA GAG CTT GGT TTG ACT TAT CTG CAC CTG ATG CCG CTG TTT			1388
Tyr Phe Gln Glu Leu Gly Leu Thr Tyr Leu His Leu Met Pro Leu Phe			
130	135	140	
AAA TGC CCT GAA GGC AAA AGC GAC GGC GGC TAT GCG GTC AGC AGC TAC			1436
Lys Cys Pro Glu Gly Lys Ser Asp Gly Gly Tyr Ala Val Ser Ser Tyr			
145	150	155	160
CGC GAT GTC AAT CCG GCA CTG GGC ACA ATA GGC GAC TTG CGC GAA GTC			1484
Arg Asp Val Asn Pro Ala Leu Gly Thr Ile Gly Asp Leu Arg Glu Val			
165	170	175	
ATT GCT GCG CTG CAC GAA GCC GGC ATT TCC GCC GTC GTC GAT TTT ATC			1532
Ile Ala Ala Leu His Ala Gly Ile Ser Ala Val Val Asp Phe Ile			
180	185	190	
TTC AAC CAC ACC TCC AAC GAA CAC GAA TGG GCG CAA CGC TGC GCC GCC			1580
Phe Asn His Thr Ser Asn Glu His Glu Trp Ala Gln Arg Cys Ala Ala			
195	200	205	
GGC GAC CCG CTT TTC GAC AAT TTC TAC TAT ATT TTC CCC GAC CGC CGG			1628
Gly Asp Pro Leu Phe Asp Asn Phe Tyr Tyr Ile Phe Pro Asp Arg Arg			
210	215	220	
ATG CCC GAC CAA TAC GAC CGC ACC CTG CGC GAA ATC TTC CCC GAC CAG			1676
Met Pro Asp Gln Tyr Asp Arg Thr Leu Arg Glu Ile Phe Pro Asp Gln			
225	230	235	240
CAC CCG GGC TTC TCG CAA CTG GAA GAC GGA CGC TGG GTG TGG ACG			1724
His Pro Gly Gly Phe Ser Gln Leu Glu Asp Gly Arg Trp Val Trp Thr			
245	250	255	
ACC TTC AAT TCC TTC CAA TGG GAC TTG AAT TAC AGC AAC CCG TGG GTA			1772
Thr Phe Asn Ser Phe Gln Trp Asp Leu Asn Tyr Ser Asn Pro Trp Val			
260	265	270	

TTC CGC GCA ATG GCG GGC GAA ATG CTG TTC CTT GCC AAC TTG GGC GTT Phe Arg Ala Met Ala Gly Glu Met Leu Phe Leu Ala Asn Leu Gly Val 275 280 285	1820
GAC ATC CTG CGT ATG GAT GCG GTT GCC TTT ATT TGG AAA CAA ATG GGG Asp Ile Leu Arg Met Asp Ala Val Ala Phe Ile Trp Lys Gln Met Gly 290 295 300	1868
ACA AGC TGC GAA AAC CTG CCG CAG GCG CAC GCC CTC ATC CGC GCG TTC Thr Ser Cys Glu Asn Leu Pro Gln Ala His Ala Leu Ile Arg Ala Phe 305 310 315 320	1916
AAT GCC GTT ATG CGT ATT GCC GCG CCC GCC GTG TTC TTC AAA TCC GAA Asn Ala Val Met Arg Ile Ala Ala Pro Ala Val Phe Phe Lys Ser Glu 325 330 335	1964
GCC ATC GTC CAC CCC GAC CAA GTC GTC CAA TAC ATC GGG CAG GAC GAA Ala Ile Val His Pro Asp Gln Val Val Gln Tyr Ile Gly Gln Asp Glu 340 345 350	2012
TGC CAA ATC GGT TAC AAC CCC CTG CAA ATG GCA TTG TTG TGG AAC ACC Cys Gln Ile Gly Tyr Asn Pro Leu Gln Met Ala Leu Leu Trp Asn Thr 355 360 365	2060
CTT GCC ACG CGC GAA GTC AAC CTG CTC CAT CAG GCG CTG ACC TAC CGC Leu Ala Thr Arg Glu Val Asn Leu Leu His Gln Ala Leu Thr Tyr Arg 370 375 380	2108
CAC AAC CTG CCC GAG CAT ACC GCC TGG GTC AAC TAC GTC CGC AGC CAC His Asn Leu Pro Glu His Thr Ala Trp Val Asn Tyr Val Arg Ser His 385 390 395 400	2156
GAC GAC ATC GGC TGG ACG TTT GCC GAT GAA GAC GCG GCA TAT CTG GGC Asp Asp Ile Gly Trp Thr Phe Ala Asp Glu Asp Ala Ala Tyr Leu Gly 405 410 415	2204
ATA AGC GGC TAC GAC CAC CGC CAA TTC CTC AAC CGC TTC TTC GTC AAC Ile Ser Gly Tyr Asp His Arg Gln Phe Leu Asn Arg Phe Phe Val Asn 420 425 430	2252
CGT TTC GAC GGC AGC TTC GCT CGT GGC GTA CCG TTC CAA TAC AAC CCA Arg Phe Asp Gly Ser Phe Ala Arg Gly Val Pro Phe Gln Tyr Asn Pro 435 440 445	2300
AGC ACA GGC GAC TGC CGT GTC AGT GGT ACA GCA GCG GCA TTG GTC GGC Ser Thr Gly Asp Cys Arg Val Ser Gly Thr Ala Ala Leu Val Gly 450 455 460	2348
TTG GCG CAA GAC GAT CCC CAC GCC GTT GAC CGC ATC AAA CTC TTG TAC Leu Ala Gln Asp Asp Pro His Ala Val Asp Arg Ile Lys Leu Leu Tyr 465 470 475 480	2396
AGC ATT GCT TTG AGT ACC GGC GGT CTG CCG CTG ATT TAC CTA GGC GAC Ser Ile Ala Leu Ser Thr Gly Gly Leu Pro Leu Ile Tyr Leu Gly Asp 485 490 495	2444
GAA GTG GGT ACG CTC AAT GAC GAC GAC TGG TCG CAA GAC AGC AAT AAG Glu Val Gly Thr Leu Asn Asp Asp Trp Ser Gln Asp Ser Asn Lys 500 505 510	2492

AGC GAC GAC AGC CGT TGG GCG CAC CGT CCG CGC TAC AAC GAA GCC CTG Ser Asp Asp Ser Arg Trp Ala His Arg Pro Arg Tyr Asn Glu Ala Leu 515 520 525	2540
TAC GCG CAA CGC AAC GAT CCG TCG ACC GCA GCC GGG CAA ATC TAT CAG Tyr Ala Gln Arg Asn Asp Pro Ser Thr Ala Ala Gly Gln Ile Tyr Gln 530 535 540	2588
GGC TTG CGC CAT ATG ATT GCC GTC CGC CAA AGC AAT CCG CGC TTC GAC Gly Leu Arg His Met Ile Ala Val Arg Gln Ser Asn Pro Arg Phe Asp 545 550 555 560	2636
GGC GGC AGG CTG GTT ACA TTC AAC ACC AAC AAG CAC ATC ATC GGC Gly Gly Arg Leu Val Thr Phe Asn Thr Asn Asn Lys His Ile Ile Gly 565 570 575	2684
TAC ATC CGC AAC AAT GCG CTT TTG GCA TTC GGT AAC TTC AGC GAA TAT Tyr Ile Arg Asn Asn Ala Leu Leu Ala Phe Gly Asn Phe Ser Glu Tyr 580 585 590	2732
CCG CAA ACC GTT ACC GCG CAT ACC CTG CAA GCC ATG CCC TTC AAG GCG Pro Gln Thr Val Thr Ala His Thr Leu Gln Ala Met Pro Phe Lys Ala 595 600 605	2780
CAC GAC CTC ATC GGT GGC AAA ACT GTC AGC CTG AAT CAG GAT TTG ACG His Asp Leu Ile Gly Gly Lys Thr Val Ser Leu Asn Gln Asp Leu Thr 610 615 620	2828
CTT CAG CCC TAT CAG GTC ATG TGG CTC GAA ATC GCC TGA CGCACGCTTC Leu Gln Pro Tyr Gln Val Met Trp Leu Glu Ile Ala *625 630 635	2877
CCAAATGCCG TCTGAACCGT TTCAGACGGC ATTTGCG	2914

(2) INFORMATION FOR SEQ ID NO: 2:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 637 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Met Leu Thr Pro Thr Gln Gln Val Gly Leu Ile Leu Gln Tyr Leu Lys 1 5 10 15
Thr Arg Ile Leu Asp Ile Tyr Thr Pro Glu Gln Arg Ala Gly Ile Glu 20 25 30
Lys Ser Glu Asp Trp Arg Gln Phe Ser Arg Arg Met Asp Thr His Phe 35 40 45
Pro Lys Leu Met Asn Glu Leu Asp Ser Val Tyr Gly Asn Asn Glu Ala 50 55 60
Leu Leu Pro Met Leu Glu Met Leu Leu Ala Gln Ala Trp Gln Ser Tyr

65	70	75	80
Ser Gln Arg Asn Ser	Ser Leu Lys Asp	Ile Asp Ile Ala Arg	Glu Asn
85	90	95	
Asn Pro Asp Trp Ile Leu Ser Asn	Lys Gln Val Gly	Gly Val Cys Tyr	
100	105	110	
Val Asp Leu Phe Ala Gly Asp	Leu Lys Gly	Leu Lys Asp Lys	Ile Pro
115	120	125	
Tyr Phe Gln Glu Leu Gly	Leu Thr Tyr	Leu His Leu Met Pro	Leu Phe
130	135	140	
Lys Cys Pro Glu Gly Lys Ser Asp	Gly Tyr Ala Val Ser	Ser Tyr	
145	150	155	160
Arg Asp Val Asn Pro Ala Leu Gly	Thr Ile Gly Asp	Leu Arg Glu Val	
165	170	175	
Ile Ala Ala Leu His Glu Ala Gly	Ile Ser Ala Val Val	Asp Phe Ile	
180	185	190	
Phe Asn His Thr Ser Asn Glu His	Glu Trp Ala Gln Arg	Cys Ala Ala	
195	200	205	
Gly Asp Pro Leu Phe Asp Asn Phe	Tyr Tyr Ile Phe	Pro Asp Arg Arg	
210	215	220	
Met Pro Asp Gln Tyr Asp Arg Thr	Leu Arg Glu Ile Phe	Pro Asp Gln	
225	230	235	240
His Pro Gly Gly Phe Ser Gln	Leu Glu Asp Gly Arg	Trp Val Trp Thr	
245	250	255	
Thr Phe Asn Ser Phe Gln Trp Asp	Leu Asn Tyr Ser Asn	Pro Trp Val	
260	265	270	
Phe Arg Ala Met Ala Gly Glu Met	Leu Phe Leu Ala Asn	Leu Gly Val	
275	280	285	
Asp Ile Leu Arg Met Asp Ala Val	Ala Phe Ile Trp	Lys Gln Met Gly	
290	295	300	
Thr Ser Cys Glu Asn Leu Pro Gln	Ala His Ala Leu	Ile Arg Ala Phe	
305	310	315	320
Asn Ala Val Met Arg Ile Ala Ala	Pro Ala Val Phe	Phe Lys Ser Glu	
325	330	335	
Ala Ile Val His Pro Asp Gln Val	Val Gln Tyr Ile Gly	Gln Asp Glu	
340	345	350	
Cys Gln Ile Gly Tyr Asn Pro Leu	Gln Met Ala Leu	Leu Trp Asn Thr	
355	360	365	
Leu Ala Thr Arg Glu Val Asn	Leu Leu His Gln Ala	Leu Thr Tyr Arg	
370	375	380	

His Asn Leu Pro Glu His Thr Ala Trp Val Asn Tyr Val Arg Ser His
 385 390 395 400

Asp Asp Ile Gly Trp Thr Phe Ala Asp Glu Asp Ala Ala Tyr Leu Gly
 405 410 415

Ile Ser Gly Tyr Asp His Arg Gln Phe Leu Asn Arg Phe Phe Val Asn
 420 425 430

Arg Phe Asp Gly Ser Phe Ala Arg Gly Val Pro Phe Gln Tyr Asn Pro
 435 440 445

Ser Thr Gly Asp Cys Arg Val Ser Gly Thr Ala Ala Ala Leu Val Gly
 450 455 460

Leu Ala Gln Asp Asp Pro His Ala Val Asp Arg Ile Lys Leu Leu Tyr
 465 470 475 480

Ser Ile Ala Leu Ser Thr Gly Gly Leu Pro Leu Ile Tyr Leu Gly Asp
 485 490 495

Glu Val Gly Thr Leu Asn Asp Asp Trp Ser Gln Asp Ser Asn Lys
 500 505 510

Ser Asp Asp Ser Arg Trp Ala His Arg Pro Arg Tyr Asn Glu Ala Leu
 515 520 525

Tyr Ala Gln Arg Asn Asp Pro Ser Thr Ala Ala Gly Gln Ile Tyr Gln
 530 535 540

Gly Leu Arg His Met Ile Ala Val Arg Gln Ser Asn Pro Arg Phe Asp
 545 550 555 560

Gly Gly Arg Leu Val Thr Phe Asn Thr Asn Asn Lys His Ile Ile Gly
 565 570 575

Tyr Ile Arg Asn Asn Ala Leu Leu Ala Phe Gly Asn Phe Ser Glu Tyr
 580 585 590

Pro Gln Thr Val Thr Ala His Thr Leu Gln Ala Met Pro Phe Lys Ala
 595 600 605

His Asp Leu Ile Gly Gly Lys Thr Val Ser Leu Asn Gln Asp Leu Thr
 610 615 620

Leu Gln Pro Tyr Gln Val Met Trp Leu Glu Ile Ala *
 625 630 635

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "oligonucleotide"

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Neisseria polysaccharea

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

CTCACCATGG GCATCTTGGG CATC

24

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "oligonucleotide"

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Neisseria polysaccharea

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

CTGCCATGGT TCAGACGGCA TTTGG

25